IMPROVED ESTIMATION OF A PATTERNED COVARIANCE MATRIX

BY

DIPAK K. DEY and ALAN E. GELFAND

TECHNICAL REPORT NO. 431

JULY 18, 1990

Prepared Under Contract

N00014-89-J-1627 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

ABSTRACT

Suppose a random vector X has a multinormal distribution with covariance matrix Σ of the form $\Sigma = \sum_{i=1}^{k} \theta_i M_i$, where the M_i's form a known complete orthogonal set and θ_i's are the distinct unknown eigenvalues of Σ. The problem of estimation of Σ is considered under several plausible loss functions. The approach is to establish a duality relationship: estimation of the patterned covariance matrix Σ is dual to simultaneous estimation of scale parameters of independent chi-square distributions. This duality allows simple estimators which, for example, improved upon the MLE of Σ. It also allows improved estimation of trΣ. Examples are given in the case when Σ has equicorrelated structure.

1. INTRODUCTION AND SUMMARY

Recently there has been considerable interest in the estimation of the covariance matrix of a multivariate normal distribution. This problem is addressed extensively in Stein (1975, 1977), Olkin and Selliah (1977), Haff (1977, 1979, 1982), and Dey and Srinivasan (1985, 1986) under plausible loss functions. However, there is no work of our kind available in the literature when the covariance matrix has an assumed structure.

Suppose a random vector X has a multinormal distribution with mean zero and covariance matrix Σ, which has the form

$$\Sigma = \sum_{i=1}^{k} \theta_i M_i \tag{1.1}$$

where the θ_i's are the distinct but unknown eigenvalues of Σ and the M_i's are a known complete orthogonal set of projection matrices. Such a structure for Σ arises in many practical

situations. A familiar example is the equicorrelated case, that is, $\Sigma = \sigma^2[(1-\rho)I_p + \rho J_p]$ where I_p is the identity matrix and J_p is a p x p matrix of 1's. This is often referred to as intraclass correlation structure. More generally, patterned covariance matrices of the form (1.1) arise naturally in variance component models. See Albert (1976) for details.

From the classical viewpoint one would estimate Σ by obtaining the maximum likelihood estimates of the θ_i's using the normality of X. In this paper, however, we take a decision theoretic approach for the estimation of Σ using the following loss structures:

$$L_q(\hat{\Sigma},\Sigma) = \mathrm{tr}(\hat{\Sigma}-\Sigma)^2 \tag{1.2}$$

and

$$L_e(\hat{\Sigma},\Sigma) = \mathrm{tr}\hat{\Sigma}\Sigma^{-1} - \log|\hat{\Sigma}\Sigma^{-1}| - p \tag{1.3}$$

The loss (1.2) is the usual extension of squared error loss (SEL) and the loss (1.3) is based on entropy measure of distance. Under these losses the MLE is inadmissible and substantial improvement is available (see Table 1). We may show that for estimators of the form $\sum_{i=1}^{k} \theta_i M_i$, these losses become, respectively,

$$L(\hat{\theta},\theta) = \sum_{i=1}^{k} p_i(\hat{\theta}_i-\theta_i)^2, \tag{1.4}$$

$$L(\hat{\theta},\theta) = \sum_{i=1}^{k} p_i[\hat{\theta}_i/\theta_i - \log(\hat{\theta}_i/\theta_i) - 1] \tag{1.5}$$

where $p_i = \mathrm{rank}(M_i)$, $\theta = (\theta_1, \ldots, \theta_k)$ and $\sum_{i=1}^{k} p_i = p$.

In addition, we note the following result essentially given in Albert (1976).

Theorem 1.1. Suppose $X \sim N_p(0,\Sigma)$. Define $Q_i = X'M_iX$, where $I_p = \sum_{i=1}^{k} M_i$, M_i being orthogonal projection matrices free from θ_i's, having $\text{rank}(M_i) = p_i$, $i=1, \ldots, k(\leq p)$. Then a necessary and sufficient condition for (1) $Q_i \sim \theta_i\chi^2_{p_i}$, $i=1, \ldots, k$, and (2) Q_i's mutually independent is $\Sigma = \sum_{i=1}^{k} \theta_i M_i$.

The equivalence of (1.2) and (1.4) and of (1.3) and (1.5) for estimators of the form $\sum_{i=1}^{k} \hat{\theta}_i M_i$ along with Theorem 1.1 establishes the following duality. Estimation of the patterned covariance matrix Σ in (1.1) under loss (1.2) ((1.3)) is dual to simultaneous estimation of the scale parameters of independent chi-square random variables under loss (1.4) ((1.5)).

It is to be noted that in the decomposition of $\Sigma = \sum_{i=1}^{k} \theta_i M_i$, the θ_i's are the distinct eigenvalues of Σ with multiplicity p_i. Thus, for example, in the equicorrelated model

$$\Sigma = \sigma^2[(1-\rho)I_p + \rho J_p],\ \theta_1 = \sigma^2(1-\rho),\ \theta_2 = \sigma^2\{1 + (p-1)\rho\},$$

$$M_1 = I_p - p^{-1}J_p,\ M_2 = p^{-1}J_p,\ p_1 = p - 1, \text{ and } p_2 = 1.$$

In Section 2, we study the estimation of Σ under loss (1.2) and also $\text{tr}\Sigma$ under SEL. We illustrate for Σ with equicorrelated structure. In this case, improved estimation of σ^2 is discussed. Additionally, improved estimation of $\rho\sigma^2$ is also considered.

Section 3 is devoted to estimation of Σ under loss (1.3). Finally, in Section 4, some encouraging numerical results are given for the equicorrelated model.

2. IMPROVED ESTIMATION UNDER L_q LOSS

2.1. Estimation of Σ^s

Here we assume the conditions of Theorem 1.1 and generalize slightly the above discussion to estimation of $\Sigma^s = \sum_{i=1}^{k} \theta_i^s M_i$ under the loss

$$L_q(\hat{\Sigma}^s, \Sigma^s) = \mathrm{tr}(\hat{\Sigma}^s - \Sigma^s)^2, \tag{2.1}$$

which, using estimators of the form $\sum_{i=1}^{k} \hat{\theta}_i^s M_i$, is dual to the estimation of $\theta^s = (\theta_1^s, \ldots, \theta_k^s)$ under the loss

$$L(\hat{\theta}^s, \theta^s) = \sum_{i=1}^{k} p_i(\hat{\theta}_i^s - \theta_i^s)^2. \tag{2.2}$$

For the estimation of Σ, s = 1; for the estimation of the precision matrix Σ^{-1}, s = -1. Let $Q = (Q_1, \ldots, Q_k)$ where $Q_i \sim \theta_i \chi^2_{p_i}$ and are independent. Suppose $\delta^0(Q)$ is an estimator of θ^s given componentwise as $\delta_i^0(Q) = a_i Q_i^s$, $i = 1, \ldots, k$. For example, $a_i = p_i^{-s}$ gives the MLE of θ_i^s.

Now define

(1) $r_{i,\alpha,\beta} = E(Q_i^\alpha | \theta_i = 1)/E(Q_i^\beta | \theta_i = 1)$

$$= 2^{\alpha-\beta}\, \Gamma((p_i + 2\alpha)/2)/\Gamma((p_i + 2\beta)/2).$$

(2) For $\alpha = (\alpha_1, \ldots, \alpha_k)$, $\nu_\alpha = \prod_{i=1}^{k} E(Q_i^{\alpha_i} | \theta_i = 1)$

$$= \prod_{i=1}^{k} 2^{\alpha_i} \Gamma((p_i + 2\alpha_i)/2)/\Gamma(p_i/2).$$

Note that $a_i = r_{i,s,2s}$ gives the best invariant estimator of θ_i^s under squared error loss.

Thus $\hat{\boldsymbol{\Sigma}}^s_M = \Sigma p_i^{-s} Q_i^s$ is the maximum likelihood estimator of $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\Sigma}}^s_0 = \Sigma r_{i,s,2s}\, Q_i^s$ is the estimator obtained by combining the best invariant estimators of θ_i^s. The following lemma shows that $\hat{\boldsymbol{\Sigma}}^s_C$ dominates $\hat{\boldsymbol{\Sigma}}^s_M$ under the risk criterion.

Lemma 2.1. $R(\hat{\boldsymbol{\Sigma}}^s_0, \boldsymbol{\Sigma}^s) \leq R(\hat{\boldsymbol{\Sigma}}^s_M, \boldsymbol{\Sigma}^s)\ \forall\ \boldsymbol{\Sigma}$.

Proof. Immediate from the duality (2.1), (2.2) and the fact that $r_{i,s,2s} Q_i^s$ dominates $p_i^{-s} Q_i^s$ in estimating θ_i^s.

In view of Lemma 2.1, it is sufficient to find estimators which improve $\hat{\boldsymbol{\Sigma}}^s_0$. We have the following theorem whose proof is a special case of a result in Dey and Gelfand (1987).

Theorem 2.1. Consider the estimator $\delta(Q) = (\delta_1(Q), \ldots, \delta_k(Q))$ given componentwise as

$$\delta_i(Q) = \delta_i^0(Q) = b\Big(\prod_{j=1}^{k} Q_j\Big)^{s/k}, \quad i = 1, \ldots, k(\geq 2), \tag{2.4}$$

where $\delta_i^0(Q) = r_{i,s,2s} Q_i^s$. Then provided all expectations exist, $\delta(Q)$ dominates $\delta^0(Q)$ under loss (2.2) if

$$0 < b < 2\nu_{s/k} d^{(1)} / \nu_{2s/k} d^{(2)}$$

where $d^{(1)} = \min(p_i |d_i|)$ and $d^{(2)} = \max(p_i)$ with

$$d_i = r_{i,s,2s} r_{i,(k+1)s/k,s/k} - 1, \quad i = 1, \ldots, k.$$

In view of Theorem 2.1, it follows that under the loss (2.1), an improved estimator of $\boldsymbol{\Sigma}^s$ is given as

$$\hat{\boldsymbol{\Sigma}}^s = \sum_{i=1}^{k} \delta_i(Q) M_i = \hat{\boldsymbol{\Sigma}}^s_0 + b\Big(\prod_{j=1}^{k} Q_j\Big)^{s/k} I_p. \tag{2.5}$$

Remark 2.1. Theorem 2.1 requires only that Q_i follow a distribution with θ_i as scale parameter. In this setting Dey and Gelfand

(1987) offer more general results than Theorem 2.1 including estimators which provide maximum improvement along a ray determined by a specified vector θ_0. These estimators shrink (expand) differently on each coordinate. Using the fact that the Q_i are distributed as multiples of chi-square random variables enables Klonecki and Zontek (1987) to provide necessary and sufficient conditions for the existence of an estimator of the form (2.4) to dominate δ^0 given componentwise as $\delta_i^0 = a_i Q_i^s$ for any specified a_i, $i = 1,\ldots,k$.

Example: Suppose Σ has equicorrelated structure. In this case $\theta_1 = \sigma^2(1-\rho)$, $\sigma_2 = \sigma^2\{1+(p-1)\rho\}$, are the distinct eigenvalues of Σ. The best invariant estimate of θ_i is $\delta_i^0(Q) = (p_i+2)^{-1} Q_i$ with $p_1 = p-1$ and $p_2 = 1$, a class of improved estimators of $\theta = (\theta_1,\theta_2)$ is given componentwise as

$$\delta_i(Q) = (p_i + 2)^{-1} Q_i + b\left(\prod_{j=1}^{2} Q_j\right)^{\frac{1}{2}}, \quad i = 1,2, \tag{2.6}$$

where $0 < b < 2\nu_{\frac{1}{2}} d^{(1)}/\nu_1 d^{(2)}$ and the upper bound on b simplifies to $2\Gamma(p/2)/3(p-1)\Gamma((p+1)/2)\sqrt{\pi}$.

The corresponding improved estimator of Σ has the simple form

$$\hat{\Sigma} = \hat{\Sigma}_0 + b(Q_1Q_2)^{\frac{1}{2}} I_p. \tag{2.7}$$

Remark 2.2. Improved estimation of $\theta^{-1} = (\theta_1^{-1},\ldots,\theta_k^{-1})$, hence of Σ^{-1}, follows directly from Theorem 2.1. We only need the existence of appropriate reciprocal moments of the Q_i. Unfortunately, in the equicorrelated case $EQ_i^{-\alpha}$ does not exist for $\alpha \geq 1$ and, hence, our approach does not provide a dominating estimator.

2.2. Estimation of trace of Σ

Consider now estimation of the trace of Σ under the SEL given as

$$L(a,\mathrm{tr}\Sigma) = (a - \mathrm{tr}\Sigma)^2. \tag{2.8}$$

Since $\mathrm{tr}\Sigma = \sum_{i=1}^{k} p_i\theta_i$ our duality converts estimation of the trace to estimation of a linear combination of chi-square scale parameters.

The following theorem gives a class of admissible estimators of $\sum_{i=1}^{k} \ell_i\theta_i$ if $\ell_i > 0$.

Theorem 2.2. If $\alpha_i > 0$, $i = 1,\ldots,k$, known then under the loss (2.8)

$$\delta_\alpha(Q) = \frac{\sum_{i=1}^{k} \ell_i\alpha_i}{p+2} \sum_{i=1}^{k} Q_i/\alpha_i \tag{2.9}$$

is admissible for $\sum_{i=1}^{k} \ell_i\theta_i$.

Proof. Consider the subset of the parameter space

$$C = \{(\theta_1,\ldots,\theta_k) : \theta_i = \alpha_i\theta,\ \alpha_i > 0,\ i = 1,\ldots,k\}.$$

Then $Q_i/\alpha_i \sim \theta\chi^2_{p_i}$, $i = 1,\ldots,k$. Thus, on C, $\sum_{i=1}^{k} Q_i/\alpha_i$ is sufficient for θ and $\sum_{i=1}^{k} Q_i/\alpha_i \sim \theta\chi^2_p$. Thus, by a theorem of Karlin (1958), $\dfrac{\sum_{i=1}^{k} Q_i/\alpha_i}{p+2}$ is admissible for θ under SEL. Then on C, $\delta_\alpha(Q)$ $= \{(\sum_{i=1}^{k} \ell_i\alpha_i) \sum_{i=1}^{k} Q_i/\alpha_i\}/(p+2)$ is admissible for $\sum_{i=1}^{k} \ell_i\alpha_i\theta = \sum_{i=1}^{k} \ell_i\theta_i$.

Suppose $\delta_\alpha(Q)$ is inadmissible for $\sum_{i=1}^{k} \ell_i\theta_i$, Then there exists $\delta^*_\alpha(Q)$ which dominates $\delta_\alpha(Q)$. But $\delta_\alpha(Q)$ being admissible on C implies $\delta^*_\alpha(Q) = \delta_\alpha(Q)$ a.s.

Note that since $p_i > 0$ in the expression for $\mathrm{tr}\ddagger$ Theorem 2.2 applies.

Remark 2.3. Consider the equicorrelated structure. As a special case, $\alpha_i = 1$ gives $\frac{p}{p+2} X'X$ admissible for trace $\ddagger$ and hence $X'X/(p+2)$ admissible for σ^2. Similarly $\alpha_i = p_i/(p_i+2)$ implies $\frac{\sum_{i=1}^k p_i^2/(p_i+2)}{p+2} \sum_{i=1}^k p_i Q_i/(p_i+2)$ is admissible for σ^2, i.e., an appropriate linear combination of the componentwise best invariant estimator is admissible.

Now we will demonstrate a general method for improving on a linear estimator of a linear combination. The improved estimators are nonlinear, and may shrink or expand the given linear estimator. Work of Das Gupta (1986), Dey and Gelfand (1987), and Klonecki and Zontek (1987) is relevant here. A general result is:

Theorem 2.3. Provided expectations exist, an estimator of $\mathrm{tr}\ddagger$ of the form $\delta_0 = \sum_{i=1}^k \ell_i Q_i$ ($\ell_i = 1$ yields $\sum_{i=1}^k Q_i = X'X$ the MLE, which is also UMVUE) is dominated by

$$\delta_{r,c} = \delta_0 + c \prod_{j=1}^k Q_j^{r_j} \tag{2.10}$$

where $r_j \geq 0$, $\sum_{j=1}^k r_j = 1$, if and only if either

(i) $d_{(1)} \geq 0$, $r_i = 0$ if $d_i = 0$ and $c > 0$ sufficiently small,

(ii) $d_{(k)} \leq 0$, $r_i = 0$ if $d_i = 0$ and $c < 0$ sufficiently large,

where $d_i = (1-\ell_i)p_i/2 - \ell_i r_i$, $i = 1,\ldots,k$ and $d_{(1)} = \min d_i$, $d_{(k)} = \max d_i$.

Proof. The risk difference between (2.10) and δ_0 is

$$\Delta(\theta) = c^2 \nu_{2r} \prod_{j=1}^k \theta_j^{2r_j} - c\nu_r \prod_{j=1}^k \theta_j^{r_j} \sum_{j=1}^k \theta_j d_j .$$

Suppose (i) holds (the proof for (ii) is similar) and that all j such that $r_j = d_j = 0$ have been deleted in $\Delta(\theta)$. Then

$$\Delta(\theta) = c\nu_{2r} \prod_{i=1}^{k} \theta_j^{2r_j}\left[c - \frac{\nu_r}{\nu_{2r}} \frac{\sum_{j=1}^{k} \theta_j d_j}{\prod_{j=1}^{k} \theta_j^{r_j}}\right].$$

But $\Sigma\theta_j d_j/\Pi\theta_j^{r_j} \geq \Pi(d_j/r_j)^{r_j}$. Hence, $\Delta(\theta) \leq 0$, $\forall\theta$, if c is positive and sufficiently small.

Remark 2.4. In particular, when $\ell_i = 1$ any set of nonnegative r_i such that $\Sigma r_i = 1$ and at least two r_i differ from zero will work. Here $c < 0$ so that the dominating estimator is a shrinker. Since $\mathrm{tr}\Sigma > 0$, $\delta^+_{r,c} = \max(\delta_{r,c}, 0)$ will dominate $\delta_{r,c}$ (using a lemma of Klotz, Milton and Zacks (1969), p. 1394).

Example. Again consider the equicorrelated structure. Clearly $\mathrm{tr}\Sigma = p\sigma^2$. Thus, using (2.10), we can explicitly dominate $X'X$ in estimating $p\sigma^2$, hence $X'X/p$ in estimating σ^2 by nonlinear estimators. For instance, the estimator

$$\delta^* = X'X/p + b(Q_1Q_2)^{\frac{1}{2}} \tag{2.11}$$

dominates the MLE under loss function $(\delta - \sigma^2)^2$ if

$$-\,4\Gamma(p/2)/p\Gamma(\tfrac{p+1}{2})\,\sqrt{\pi} < b < 0.$$

Remark 2.5. If we attempt to apply Theorem 2.3 to linear estimators of the form (2.9), we will discover that all d_i's are equal to zero. In particular, in the above example, we cannot dominate $X'X/(p+2)$ in estimating σ^2.

Continuing with our example, suppose $\rho \geq 0$ and consider estimation of $\rho\sigma^2$ which may be viewed as variance component. (See Gelfand and Dey

(1988) for more detailed discussion of improved estimation of variance components.) Since $\rho\sigma^2 = (\theta_2-\theta_1)/p$, we consider the estimator

$$\delta^0 = p^{-1}(a_2Q_2 - a_1Q_1)$$

where $a_i = (p_i + 2\varepsilon_i)^{-1}$, $i = 1,2$, with $0 \le \varepsilon_i \le 1$. For example, taking $\varepsilon_i = 0$, $i = 1,2$, δ^0 becomes the MLE of $\rho\sigma^2$ and taking $\varepsilon_i = 1$, δ_0 is formed from the best invariant estimator of θ_i, under SEL, $i = 1,2$. A class of improved estimators of $\rho\sigma^2$, is given as

$$\delta = \delta^0 + b(a_1Q_1)^r(a_2Q_2)^{1-r}, \tag{2.12}$$

using Theorem 2.3, provided $r > 0$ can be chosen such that either (i) $r < \min(\varepsilon_1, 1 - \varepsilon_2)$ whence b must be positive and sufficiently small or (ii) $r > \max(\varepsilon_1, 1 - \varepsilon_2)$ whence $-b$ must be positive and small. In fact, we would use δ^+.

Remark 2.6. Note that Theorem 2.2 does not provide admissible estimators of $\rho\sigma^2$ since $\ell_i < 0$.

3. IMPROVED ESTIMATION UNDER L_c LOSS

In this section we consider the estimation of patterned Σ under loss (1.3). Using the aforementioned duality for estimators of the form $\sum_{i=1}^{k} \hat{\theta}_i M_i$, we convert this problem to simultaneous estimation of the eigenvalues $\theta = (\theta_1,\ldots,\theta_k)$ of Σ under the loss (1.5). In fact, our results can be extended to the estimation of $\theta^s = (\theta_1^s,\ldots,\theta_k^s)$ and hence the estimation of Σ^s. However, it is not clear how to apply loss (1.3) in estimating $\mathrm{tr}\Sigma$ which is not a scale parameter.

Our approach is that of Berger (1980) and Dey, Ghosh and Srinivasan (1987). Assume Q_i are independent Gamma(α_i,η_i) $(\alpha_i > 0,\ \eta_i > 0)$ random variables, having density

$$f(Q_i|\eta_i) = \eta_i^{\alpha_i} Q_i^{\alpha_i - 1} e^{-Q_i\eta_i}/\Gamma(\alpha_i). \tag{3.1}$$

In our case, we have $\alpha_i = p_i/2$, $\eta_i = (2\theta_i)^{-1}$, whence the loss (1.5) corresponds to simultaneous estimation of the η_i^{-1}; that is,

$$L(\delta,\eta^{-1}) = \sum_{i=1}^{k} p_i[\delta_i\eta_i - \log(\delta_i\eta_i) - 1].$$

Since the MLE of θ_i is $\delta_i^0(Q) = Q_i/p_i$, the MLE and the unbiased estimator of $\theta_i = \eta_i^{-1}/2$ is $\delta_i^0(Q) = Q_i/2\alpha_i$, $i = 1,\dots,k$. From Dey, Ghosh and Srinivasan (1987), it follows that $\delta^0(Q) = (\delta_1^0(Q),\dots,\delta_k^0(Q))$ is admissible for $k = 2$ if $\min(\alpha_1,\alpha_2) > 4$. To seek a dominating estimator, we require $k > 2$.

Now assuming the conditions in Lemma 1 of Berger (1980), it follows that if

$$\delta(Q) = \delta^0(Q) + \phi(Q) \tag{3.2}$$

is a rival estimator, the risk difference is

$$\Delta(\theta) = R(\delta,\theta) - R(\delta^0,\theta) = E_\theta\Delta_0(Q),$$

where $\Delta_0(Q)$ is the unbiased estimate of the risk difference given as

$$\Delta_0(Q) = \sum_{i=1}^{k} p_i[\phi_i^{i(1)}(Q) + (\alpha_i - 1)\phi_i(Q)/Q_i - \log(1+\alpha_i\phi_i(Q)/Q_i)]$$

with $\phi_i^{i(1)}(Q) = \partial\phi_i(Q)/\partial Q_i$. Defining $\phi_i(Q) = Q_i\psi_i(Q)$, one gets

$$\Delta_0(Q) = \sum_{i=1}^{k} p_i[Q_i\psi_i^{i(1)}(Q) + \alpha_i\psi_i(Q) - \log(1 + \alpha_i\psi_i(Q))]. \tag{3.3}$$

In order to obtain an improved estimator $\delta(Q)$, it is sufficient to find a solution $\Delta_0(Q) \le 0$ with strict inequality for some set of Q.

The following theorem gives a class of dominating shrinkage estimators.

Theorem 3.1. Suppose $S = \Sigma \log^2(Q_i/2)$. Consider an estimator $\delta(Q) = (\delta_1(Q),\ldots,\delta_k(Q))$ given componentwise as

$$\delta_i(Q) = p_i^{-1}Q_i - Q_i\tau(S)\log(Q_i/2)/2(b+S),\ i = 1,\ldots,k, \tag{3.4}$$

with $b > (5.76)(k-2)^2/p^{*2}$, where $p^* = \max p_i$ and $\tau(S)$ is a function satisfying

(i) $0 < \tau(S) < 4.8(k-2)/p^{*2}$

(ii) $\tau(S) \uparrow$ in S and

(iii) $E[\tau'(S)] < \infty$. (3.5)

Then $\delta(Q)$ dominates $\delta^0(Q)$ for $k \geq 3$, in terms of risk.

Proof. The argument is similar to that of Theorem 3.1 of Dey, Ghosh and Srinivasan (1987).

Remark 3.1. Using Theorem 3.2 and 3.3 of Dey, Ghosh and Srinivasan (1987), adaptive estimators and trimmed shrinkage estimators of θ can be obtained as well.

In concluding this section, we observe another illustration of our duality relationship. Improved estimation of patterned Σ under the scale invariant loss

$$L_I(\hat{\Sigma},\Sigma) = \mathrm{tr}(\hat{\Sigma}\Sigma^{-1} - I)^2 \tag{3.6}$$

using estimators of the form $\sum_{i=1}^{k} \hat{\theta}_i M_i$, converts to simultaneous estimation of $\theta = (\theta_1,\ldots,\theta_k)$ under loss

$$L(\hat{\theta},\theta) = \sum_{i=1}^{k} p_i(\hat{\theta}_i/\theta_i - 1)^2. \tag{3.7}$$

Again, Berger's approach yields a differential inequality (the only difference will be the presence of the weights, p_i) whose solution leads to dominating estimators similar to those in (3.4). Details are omitted.

4. NUMERICAL RESULTS

To study the performance of the MLE $\hat{\Sigma}_M = \sum_{i=1}^{2} p_i^{-1} Q_i$, $\hat{\Sigma}_0 = \sum_{i=1}^{2} (p_i + 2)^{-1} Q_i$ and $\hat{\Sigma} = \hat{\Sigma}_0 + b(Q_1 Q_2)^{\frac{1}{2}} I_p$, we calculate risks for different values of p and ρ in the equicorrelated structure. We took $\sigma^2 = 1$ and $b = \Gamma(p/2)/3(p-1)\Gamma((p+1)/2)\sqrt{\pi}$, which is the midpoint of the allowable range. We then computed the percentage improvements for selected values of p and ρ $(-(p-1)^{-1} \leq \rho \leq 1)$. The improvements of $\hat{\Sigma}$ over the MLE are substantial. While the percentage improvements in risk of $\hat{\Sigma}$ over $\hat{\Sigma}_0$ are small, the simplicity of $\hat{\Sigma}$ encourages its use.

TABLE 1

PERCENTAGE IMPROVEMENTS OVER $\hat{\Sigma}_M$ AND $\hat{\Sigma}_0$

ρ	$PI1 = \frac{R(\hat{\Sigma}_M) - R(\hat{\Sigma})}{R(\hat{\Sigma}_M)} \times 100$	$PI2 = \frac{R(\hat{\Sigma}_0) - R(\hat{\Sigma})}{R(\hat{\Sigma}_0)} \times 100$
	p = 2	
-.75	67.94	3.83
-.50	68.43	5.31
-.25	68.78	6.35
0	68.92	6.75
.25	68.78	6.35
.50	68.44	5.31
.75	67.94	3.83
	p = 6	
0	49.25	3.11
.25	63.77	2.45
.50	66.42	1.50
.75	66.85	.85
	p = 10	
0	43.50	1.87
.25	64.66	1.23
.50	66.50	.70
.75	66.75	.39

ACKNOWLEDGMENT

The authors wish to acknowledge Len Kelly for performing the computations.

AMS 1980 Subject Classifications. Primary 62F10; secondary 62C99.

BIBLIOGRAPHY

Albert, A. (1976). When is a sum of squares an analysis of variance? Ann. Statist. 4, 775-778.

Berger, J. O. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. Ann. Statist. 8, 545-571.

Das Gupta, A. (1986). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. Ann. Statist. 14, 206-219.

Dey, D. K. and Gelfand, A. E. (1987). Improved estimators in simultaneous estimation of scale parameters. Technical Report No. 399, Department of Statistics, Stanford University.

Dey, D. K., Ghosh, M., and Srinivasan, C. (1987). Simultaneous estimation of parameters under entropy loss. J. Statist. Planning and Inference 15, 347-363.

Dey, D. K. and Srinivasan, C. (1985). Estimation of covariance matrix under Stein's loss. Ann. Statist. 13, 1581-1591.

Dey, D. K. and Srinivasan, C. (1986). Trimmed minimax estimator of covariance matrix. Ann. Inst. Statist. Math. 38A, 101-108.

Gelfand, A. E. and Dey, D. K. (1988). Improved estimation of a variance component in mixed models. Unpublished manuscript.

Haff, L. R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. Ann. Statist. 7, 1264-1276.

Haff, L. R. (1982). Solutions of the Euler-Lagrange equations for certain multivariate normal estimation problems. Unpublished manuscript.

Karlin, S. (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist. 2, 406-436.

Klonecki, W. and Zontek, S. (1987). Inadmissibility results for linear simultaneous estimation in the multiparameter gamma distribution. Preprint 389, Institute of Mathematics, Polish Academy of Sciences.

Klotz, J., Milton, R., and Zacks, S. (1969). Mean square efficiency of estimator of variance components. J. Amer. Statist. Assoc. 64, 1383-1402.

Olkin, I. and Selliah, J. B. (1977). Estimating covariance in a multivariate normal distribution. Statistical Decision Theory and Related Topics II, 313-326, S. S. Gupta and D. S. Moore, Academic Press, New York.

Stein, C. (1977). Unpublished notes on estimating the covariance matrix.

Zacks, S. and Ramig, P. F. (1987). Confidence intervals for the common variance of equicorrelated normal random variables. Contributions to the Theory and Application of Statistics (A. E. Gelfand, ed.), Academic Press, New York.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 431	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Improved Estimation Of A Patterned Covariance Matrix		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dipak K. Dey and Alan E. Gelfand		8. CONTRACT OR GRANT NUMBER(s) N00014-89-J-1627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE July 18, 1990
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also supported in part by The University of Connecticut Research Foundation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

patterned covariance matrix; loss function; simultaneous estimation; equicorrelated model.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE FOLLOWING PAGE.

TECHNICAL REPORT NO. 431

20. ABSTRACT

Suppose a random vector X has a multinormal distribution with covariance matrix Σ of the form $\Sigma = \sum_{i=1}^{k} \theta_i M_i$, where the M_i's form a known complete orthogonal set and θ_i's are the distinct unknown eigenvalues of Σ. The problem of estimation of Σ is considered under several plausible loss functions. The approach is to establish a duality relationship: estimation of the patterned covariance matrix Σ is dual to simultaneous estimation of scale parameters of independent chi-square distributions. This duality allows simple estimators which, for example, improved upon the MLE of Σ. It also allows improved estimation of $\mathrm{tr}\Sigma$. Examples are given in the case when Σ has equicorrelated structure.